A Structural Theorem for Local Algorithms with Applications to Coding, Testing, and Privacy

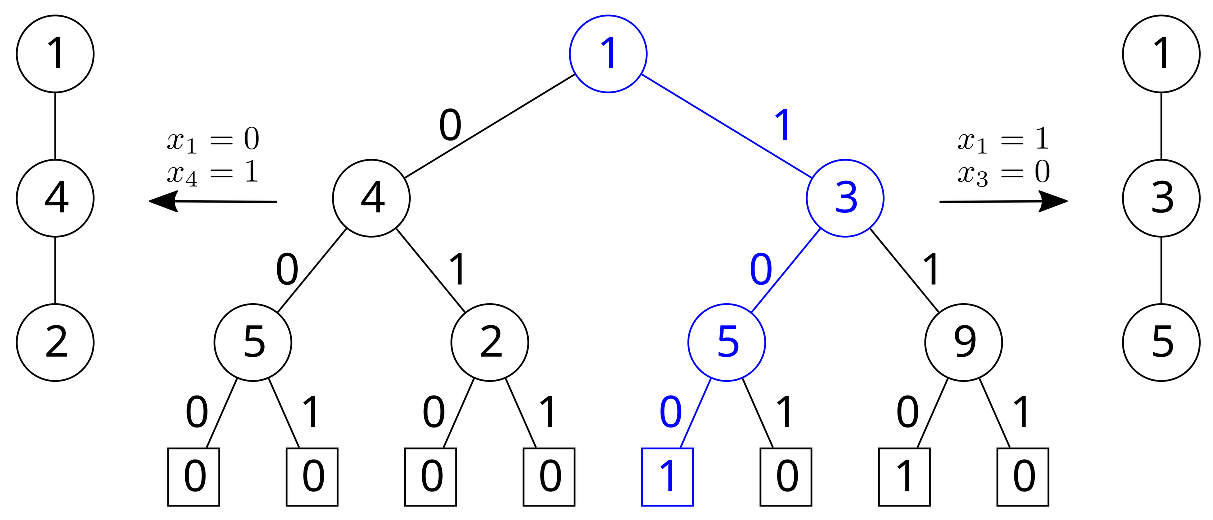
Marcel Dall'Agnol (with Tom Gur and Oded Lachish)

q-query algorithms satisfying a natural robustness condition can be made **sample-based** with sample complexity $n^{1-\tilde{\Omega}(1/q^2)}$.

This follows from a **daisy partition lemma** on set systems, bypassing an exponential blowup the general transformation incurs.

Introduction

A q-query algorithm receives $x \in \{0,1\}$ as input and inspects at most $q = O(1)$ of its coordinates to arrive at a decision. In general, queries may depend on the bits seen in past coordinates; such *adaptive* algorithms are described by decision trees:



This 3-query decision tree outputs 1 if $(x_1, x_3, x_5) = (1, 0, 0)$, when it queries the set of coordinates $\{1, 3, 5\}$. When $(x_1, x_4) = (0, 1)$ it queries $\{1, 4, 2\}$; the two other possibilities are $\{1, 4, 5\}$ and $\{1, 3, 9\}$. There is an easy way to make queries independent from the bits they return: query the entire tree, i.e., $\{1, 2, 3, 4, 5, 9\}$. The (non-adaptive) resulting algorithm makes 2^{q-1} queries, an *exponential* blowup. While 2^q is still constant, to further make the algorithm *sample-based* (i.e., query each coordinate independently with the same probability), the sample complexity becomes $n^{1-1/\exp(q)}$.

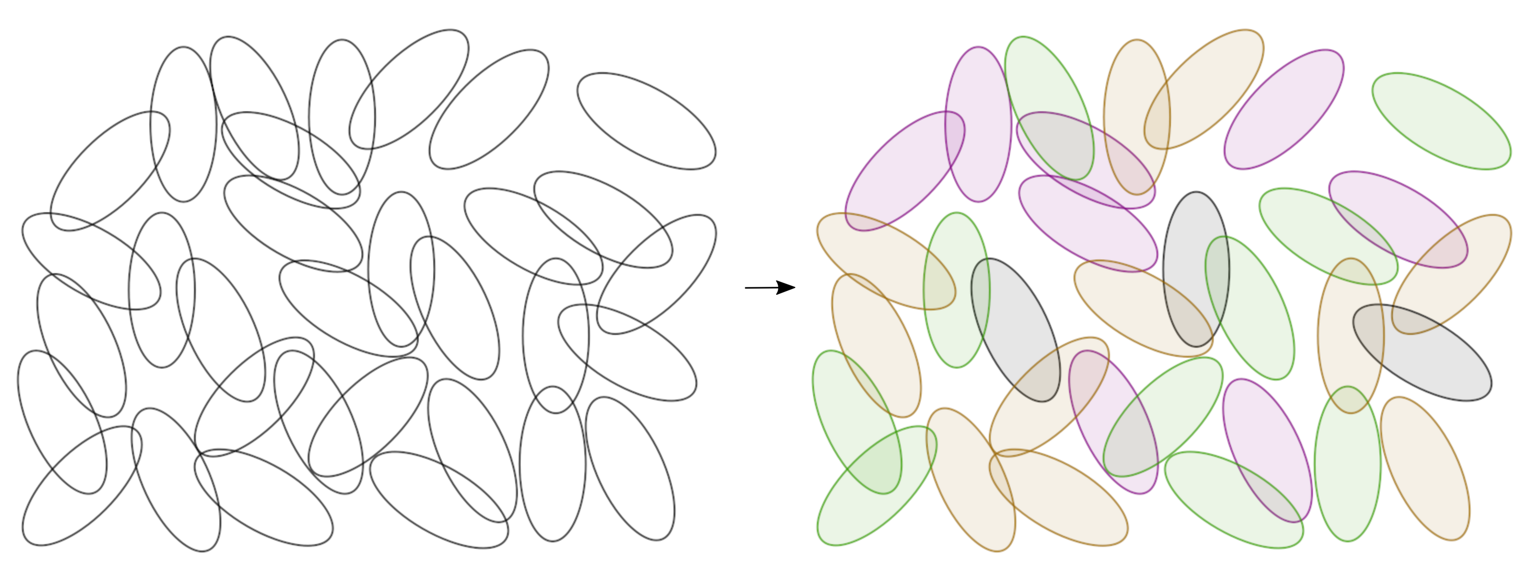
Robustness and relaxed sunflowers

We introduce a natural notion of *robustness* for query algorithms that allows to bypass the transformation above, and results in a much smaller sample complexity of $n^{1-1/\operatorname{poly}(q)}$. An algorithm A is robust at an input x if it behaves similarly for any string close to x; this is satisfied by testers, local decoders, PCPs of proximity and more.

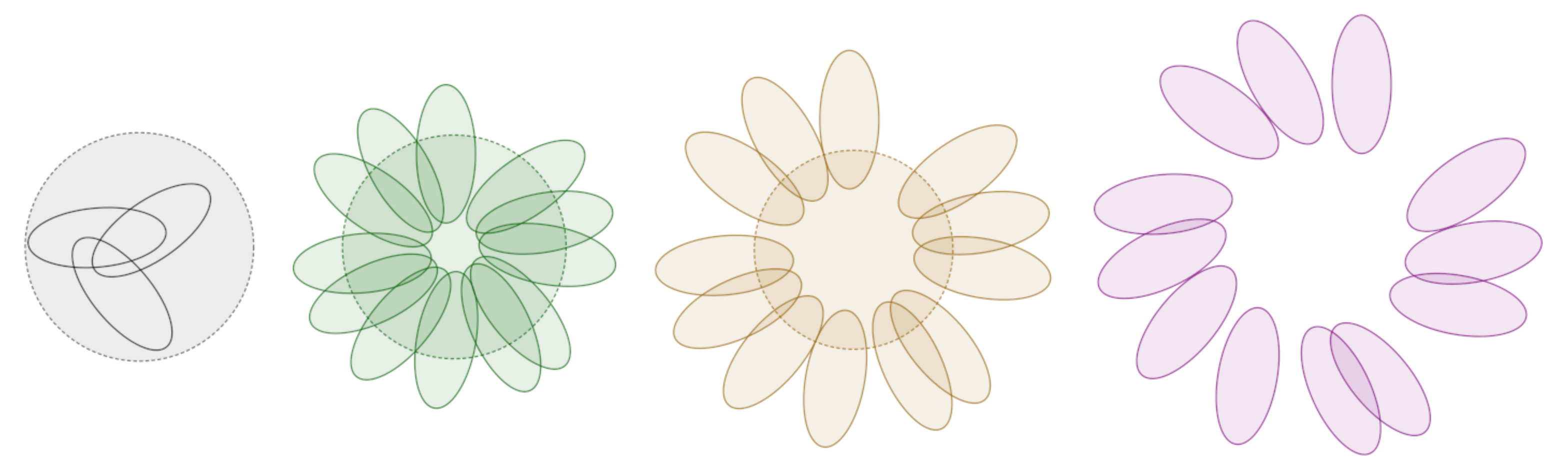
The query behaviour of robust algorithms is captured by a relaxation of the set systems known as sunflowers. We call this relaxation a *daisy*, which allows some (but not much) intersection between petals and a kernel that is not entirely contained in every set of the system.

The partition lemma

Let $\mathcal{S}$ be an arbitrary collection of q-sets of size $|\mathcal{S}| = \Theta(n)$ (e.g., the collection of sets which an algorithm may query). The *daisy partition lemma* divides $\mathcal{S}$ into a sequence of $q+1$ daisies with increasing petals and decreasing kernels.



This collection (where $q = 3$) is partitioned into *daisies* $\mathcal{D}_0$, $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$. Grouping each daisy separately yields the following:



Circles denote the *kernels* $K_0 = K_1 \subseteq K_2 \subseteq K_3 = \varnothing$, and $S \in \mathcal{D}_i$ has a *petal* $S \setminus K_i$ of size i. For $i > 0$, the i^{th} kernel has size $|K_i| = O(n^{1-i/q})$ and each petal of $\mathcal{D}_{i+1}$ intersects $t_i = O(n^{i/q})$ other petals.

Main theorem

A daisy partition of the query sets $\mathcal{S}$ of an algorithm A leads to the following sample-based algorithm B:

- Sample each coordinate $j \in [n]$ independently.
- For each daisy $\mathcal{D}_i$ and assignment $\kappa : K_i \to \{0,1\}$:
  - Consider the query sets $S \in \mathcal{D}_i$ whose petals were fully sampled. Count the number of induced assignments $a : S \to \{0,1\}$ that lead A to accept.
- Accept if any count crosses a threshold τ_i, rejecting otherwise.

While B is not always guaranteed to solve the same problem as A, we show that this is true *when A is robust*. This follows from the bounds on $|K_i|$ and t_i: the former implies that A behaves similarly for any kernel assignment, and, along with the latter, allows a union bound over all $2^{|K_i|}$ of them in a key step of the analysis.

Theorem 1. Every robust q-query algorithm can be transformed into a sample-based algorithm with sample complexity $n^{1-\Omega(1/q^2 \log^2 q)}$.

Applications

With this general transformation, we prove novel results for a number of robust algorithms. The first is an exponential improvement for the dependency on q of a lower bound for *relaxed locally decodable codes*, narrowing it down to quadratic in the known upper bound. This requires adapting the main result to *relaxed* robust algorithms.

Theorem 2. Any q-query relaxed LDC $C\colon \{0,1\}^k \to \{0,1\}^n$ with decoding radius $\Omega(1)$ must have blocklength at least $n = k^{1+\tilde{\Omega}(1/q^2)}$.

Corollary 1. Any property that is ε-testable with q queries admits a sample-based 2ε-tester with sample complexity $n^{1-\tilde{\Omega}(1/q^2)}$.

Lastly, we show that the known separation between *proofs of proximity* and testers is almost tight (also up to a quadratic factor in q).

Corollary 2. Any property that admits a MA proof of proximity of length p with query complexity q is testable with $p \cdot n^{1-\tilde{\Omega}(1/q^2)}$ queries.